2 OVERVIEW OF FINAL STATUS SURVEY DESIGN

2.1 Introduction

It is recognized that demonstrating that residual concentrations of radioactivity at a site are at very low levels in the presence of background may be a complex task involving sophisticated sampling, measurement, and statistical analysis techniques. The difficulty of the task can vary substantially depending on a number of factors, including the radionuclides in question, the background level for those and other radionuclides at the site, and the temporal and spatial variations in background at or near the site. Sufficient radiological data must be collected to characterize both the residual radioactivity at the site and the background radioactivity levels in the vicinity of the site. The number of measurements required to accomplish this task will be determined on a site-specific basis and will depend upon the nature of the facility, its size, the selection of the statistical tests used, and certain statistical parameter values that influence how compliance with radiological criteria is determined.

2.2 Final Status Survey Design

Decommissioning is defined in 10 CFR 20.1003 as removing a facility or site safely from service, and reducing residual radioactivity to a level that permits (1) release of the property for unrestricted use and termination of the license; or (2) release of the property under restricted conditions and termination of the license. A *survey unit* is a geographical area of specified size and shape for which a separate decision will be made whether or not that area meets the release criteria. This decision is made following a *final status survey* of the survey unit. Thus, a survey unit is an area for which a final status survey is designed, conducted, and results in a release decision. The objective of this report is the design of efficient final status surveys. These surveys should obtain all of the data required for making the decision, but avoid the collection and analysis of superfluous samples.

Usually there are two conditions that would lead to the determination that a particular survey unit requires further remediation before unrestricted release:

(1) If the average level of residual radioactivity within the survey unit exceeds the regulatory limit, or

(2) If there are small areas within the survey unit with elevated residual radioactivity that exceed the regulatory limit.

Sampling at discrete points within the survey unit is a simple method for determining if the first of these conditions exists. The term *sampling* is used here in its statistical sense, namely obtaining data from a subset of a population. Sampling in this sense would include both direct *in situ* measurements and the collection of physical samples for laboratory analysis.

On the other hand, sampling at discrete points within a survey unit is not a very efficient method of determining if the second condition exists. Scanning is a much better method for detecting isolated areas with elevated activity. However, scanning is generally not as sensitive as sampling.

A major component of the survey designs discussed in this report is the efficient use of sampling at distinct locations combined with scanning to accurately determine the final status of a survey unit. The statistical procedures described in this report are used to establish the number of samples taken at distinct locations needed to determine if the mean concentration in the survey unit exceeds the regulatory limit, with a specified degree of precision. Thus, these statistical procedures are as important in the planning and design of the final status survey as they are in the analysis and interpretation of the resulting data.

2.2.1 Release Criteria

In the past, release criteria have often been expressed as activity concentration limits. The criteria for license termination given in 10 CFR 20.1402 and 10 CFR 20.1403 are expressed in terms of total effective dose equivalent (TEDE). This cannot be measured directly. *Exposure pathway modeling* is used to calculate the estimated volume or surface area concentration of specific radionuclides that could result in a TEDE equal to the release criterion. This concentration is termed the *derived concentration guideline level (DCGL)*. The units for the DCGL are the same as the units for measurements performed to demonstrate compliance (e.g., Bq/kg, Bq/m^2, etc.). This allows direct comparisons between the survey results and the DCGL.

A complete discussion of DCGLs is beyond the scope of this report. There is, however, one aspect of exposure pathway modeling that bears directly on survey unit design. That is the dependence of the TEDE on the assumed area of contamination used in the exposure pathway model.

The two conditions of Section 2.2 that may cause a survey unit to fail the TEDE release criterion may have very different corresponding DCGLs because of the different size of the areas of residual radioactivity. Consequently, this report considers two distinct DCGLs:

(1) The $DCGL_W$ is derived assuming that residual radioactivity is uniformly distributed over a wide area, i.e. the entire survey unit. This can often be the default DCGL provided by an exposure pathway model.

(2) The $DCGL_{EMC}$ is derived assuming that residual radioactivity is concentrated in a much smaller area, i.e., in only a small percentage of the entire survey unit.

The $DCGL_{EMC}$ can never be less than the $DCGL_W$, but it may be significantly greater. The ratio of the $DCGL_{EMC}$ to the $DCGL_W$ defines a radionuclide specific *area factor*, F_A, such that the $DCGL_{EMC} = (F_A)\ (DCGL_W)$, when the residual radioactivity is confined to an area of size A.

Detailed procedures for developing these area factors are beyond the scope of this report. However, in the simplest case, an area factor can be determined from the ratio of the result obtained from an exposure pathway model using the entire survey unit area to the result obtained assuming the residual radioactivity is confined to a smaller area. The value of the $DCGL_{EMC}$ that is calculated for survey planning purposes is based on an area, A, determined by the spacing between adjacent sampling locations.

2.2.2 Data Interpretation

The use of the two DCGLs discussed above differs when interpreting the results of the final status survey data. The $DCGL_W$ is used to form a statistical hypothesis concerning the level of residual radioactivity that may be uniformly distributed across the survey unit. A nonparametric test is applied to the sampling data taken at distinct locations in the survey unit to determine whether this level meets the release criterion.

The $DCGL_{EMC}$, however, is used to trigger further investigation of a portion of the survey unit. Any measurement from the survey unit is considered elevated if it exceeds the $DCGL_{EMC}$. This is the *elevated measurement comparison.* The existence of an elevated measurement in a survey unit indicates the possibility of an area of residual radioactivity that may cause the dose criteria to be exceeded. The elevated measurement alone does not indicate that the survey unit fails to meet the release criterion, only that it is a possibility that must be investigated further. The $DCGL_{EMC}$ is based on the area factor used for the survey design. The area factor used in the survey design is based on the area bounded by adjacent sampling points. The actual area of elevated activity could be smaller. Thus, the area factor based on the actual area of contamination may be larger. Further investigation will usually be necessary to determine the actual extent and concentration level of a specific elevated area.

2.2.3 Survey Unit Classification

To maximize the efficiency of the final status surveys, it is clear that the greatest effort should be expended on the areas that have the highest potential for contamination. Final status survey designs depend fundamentally on the *classification* of survey units according to contamination potential. The survey unit classification determines the final status survey design and the procedures used to develop the design.

Areas that have no potential for residual contamination are classified as *non-impacted areas*. These areas have no radiological impact from site operations and are typically identified early in decommissioning. Areas with some potential for residual contamination are classified as *impacted areas*.

Impacted areas are further divided into one of three classifications:

(1) *Class 1 Areas*: Areas containing locations where, prior to remediation, the concentrations of residual radioactivity may have exceeded the $DCGL_W$.

(2) *Class 2 Areas*: Areas containing no locations where, prior to remediation, the concentrations of residual radioactivity may have exceeded the $DCGL_W$.

(3) *Class 3 Areas*: Areas with a low probability of containing any locations with residual radioactivity.

Class 1 areas have the greatest potential for contamination and therefore receive the highest degree of survey effort for the final status survey. Non-impacted areas do not receive any level of survey coverage because they have no potential for residual contamination. Impacted areas for

which there is insufficient information to justify a lower classification should be classified as Class 1.

Examples of Class 1 areas include: (1) site areas previously subjected to remedial actions, (2) locations where leaks or spills are known to have occurred, (3) former burial or disposal sites, (4) waste storage sites, and (5) areas with contaminants in discrete solid pieces of material and high specific activity.

Remediated areas are identified as Class 1 areas because the remediation process often results in less than 100% removal of the contamination. The contamination that remains on the site after remediation is often associated with relatively small areas with elevated levels of residual radioactivity. This results in a non-uniform distribution of the radionuclide and a Class 1 classification. If an area is expected to have levels of residual radioactivity below the $DCGL_W$ and was remediated for purposes of ALARA, the remediated area might be classified as Class 2 for the final status survey.

Examples of areas that might be classified as Class 2 for the final status survey include: (1) locations where radioactive materials were present in an unsealed form, (2) potentially contaminated transport routes, (3) areas downwind from stack release points, (4) upper walls and ceilings of buildings or rooms subjected to airborne radioactivity, (5) areas handling low concentrations of radioactive materials, and (6) areas on the perimeter of former contamination control areas.

To justify changing the classification from Class 1 to Class 2, there should be measurement data that provides a high degree of confidence that no individual measurement would exceed the $DCGL_W$. Other justifications for reclassifying an area as Class 2 may be appropriate, based on site-specific considerations.

Examples of areas that might be classified as Class 3 include buffer zones around Class 1 or Class 2 areas, and areas with very low potential for residual contamination but insufficient information to justify a non-impacted classification.

The number of distinct sampling locations needed to determine if a uniform level of residual radioactivity within a survey unit exists does not depend on the survey unit size. However, the sampling density within a survey unit should reflect the potential for small elevated areas of residual radioactivity. Thus, the appropriate size for survey units formed within each of the three area classifications differs. Survey units with a higher potential for residual radioactivity should be smaller. Suggested maximum areas for survey units are:

Class 1 Structures . . . . . 100 m^2 floor area
Class 1 Land areas . . . . . 2,000 m^2
Class 2 Structures . . . . . 100 to 1,000 m^2
Class 2 Land areas . . . . . 2,000 to 10,000 m^2
Class 3 Structures . . . . . no limit
Class 3 Land areas . . . . . no limit

The area of the survey unit should also be consistent with that assumed in the exposure pathway model used to calculate the $DCGL_W$. Survey units with structure surface areas less than 10 m^2 or land areas less than 100 m^2 may have unnecessarily high sampling densities, and should be avoided.

2.2.4 Final Status Survey Classification

Class 1 areas have the highest potential for containing small areas of elevated activity exceeding the release criterion. Consequently, both the number of sampling locations and the extent of scanning effort is the greatest. The final status survey is driven by the effort to provide reasonable assurance that if any areas with concentrations in excess of the $DCGL_{EMC}$ exist that then these areas will be found. Sampling is done on a systematic grid. The distance between sampling locations is made small enough that any elevated area that might be missed by sampling would be found by scanning. Scanning is performed over 100% of the survey unit. The minimum detectable concentration (MDC) of the scanning method must be lower than the $DCGL_{EMC}$.

Class 2 areas may contain residual radioactivity, but the potential for elevated areas is very small. Sampling is done on a systematic grid. The distance between samples is limited by limiting the maximum size of the survey unit. Scanning is performed systematically over the survey unit. Since Class 2 is an intermediate classification, scanning coverage may range from as little as 10% to nearly 100% of the survey unit, depending on whether the potential for an elevated area is nearer that for a Class 1 area or for a Class 3 area.

Class 3 areas should contain little, if any, residual radioactivity. There should be virtually no potential for elevated areas. Sampling is random across the survey unit, and the sample density can be very low. Scanning is limited to those parts of the survey unit where it is deemed prudent, based on the judgment of an experienced professional.

Table 2.1 summarizes the differences in the final status survey design for each of the three survey unit classifications.

Table 2.1 Final Status Survey Design Classification

Class	Sampling	Scanning
1	Systematic	100% Coverage
2	Systematic	10 – 100%
3	Random	Judgmental

2.2.5 Background

The release criteria in 10 CFR Part 20.1402 and 1403 specify a dose limit (TEDE) due to residual radioactivity that is *distinguishable from background radiation*. According to 10 CFR 20.1003, *background radiation* means radiation from cosmic sources, naturally occurring radioactive material, including radon (except as a decay product of source or special nuclear material), and global fallout as it exists in the environment from the testing of nuclear explosive devices or from nuclear accidents like Chernobyl which contribute to background radiation and are not under the control of the licensee. Background radiation does not include radiation from source, byproduct, or special nuclear materials regulated by the Commission. The term *distinguishable from background* means that the detectable concentration of a radionuclide is statistically different from the background concentration of that radionuclide in the vicinity of the site, or, in the case of structures, in similar materials using adequate measurement technology, survey and statistical techniques.

For the purposes of survey design, the method of accounting for background radiation will depend not only on the radionuclides involved, but also on the type of measurements made. For radionuclide specific measurements of radionuclides that do not appear in natural background, it is clear that no adjustments for background are needed. In some cases, a sample-specific background adjustment may be possible. For example, residual ^{238}U activity may be distinguishable from natural ^{238}U by the amount of ^{226}Ra present in a sample. In other cases, it will not be possible to make such a distinction. In particular, such a distinction will not be possible, even if the radionuclide does not appear in background, when gross activity or exposure rate measurements are used.

For the elevated measurement comparison of individual sampling results, an adjustment for background will not ordinarily be necessary, since the $DCGL_{EMC}$ is a multiple of the $DCGL_W$. For statistical testing of the results against the release criterion, however, one approach is used when the measurements represent net residual radioactivity, but a different approach is necessary when the measurements represent total radioactivity including background.

When a specific background can be established for individual samples, the results of the survey unit measurements can be compared directly to the DCGL, since each is a measurement of the residual radioactivity alone. Because only one set of measurements is involved in this comparison, the statistical test is called a *one-sample test*.

When a specific background cannot be established for individual samples, the survey unit measurements cannot be directly compared to the DCGL, since each is a measurement of the total of any residual radioactivity plus the survey unit background. In this case, the measurements in a survey unit must be compared to similar measurements in local *reference areas* that have been matched to the survey unit in terms of geological, chemical, and biological attributes, but which have not been affected by site operations. The *distribution* of the measurements in a survey unit is compared to the *distribution* of background measurements in a reference areas. Because two sets of measurements are used in making this comparison, the statistical test is called a *two-sample test*.

2.2.6 Data Variability

The ease or difficulty with which compliance may be demonstrated depends primarily on the size of the $DCGL_W$ relative to the amount of variability in the measurement data. This is commonly known as the signal-to-noise ratio. As this ratio becomes smaller, more measurement data will be needed to determine compliance with the release criterion, i.e. to extract the signal from the noise.

The variability in the measurement data is a combination of the precision of the measurement process, and the real spatial variability of the quantity being measured in the survey unit. Variability can be reduced by using more precise measurement methods, but the spatial variability remains. The mechanism by which spatial variability can be reduced is by choosing survey units that are as homogeneous as possible with respect to the expected level of residual radioactivity. This means that survey units should generally be formed from areas with similar construction, use, contamination potential, and remediation history.

If the measurement data include a background contribution, the spatial variability of background adds to the overall measurement variability. Thus, the survey units where such measurements will be used should be as homogeneous as possible with respect to expected natural background as well. Further information on natural background and its variability can be found in NUREG-1501 (August 1994).

An additional source of variability is introduced when survey unit measurement data including background are compared to measurements from a reference area. Any systematic difference in background level between the survey unit and the reference area will be indistinguishable from a difference in residual radioactivity in the two areas. This situation is not unique to decommissioning or the methodology of this report. It is always true when a background adjustment must be made using data from a location other than the sampling location, e.g. using control dosimeters at remote locations to account for background in monitoring dosimeters.

2.2.7 Reference Areas

A *reference area* (or background area) is a geographical area from which representative samples of background will be selected for comparison with samples collected in specific survey units at the remediated site. The reference area should have similar physical, chemical, radiological, and biological characteristics to the site area being remediated, but should not have been contaminated by site activities. The reference area is where background is measured and defined for the purpose of decommissioning. To minimize systematic biases in the comparison, the same sampling procedure, measurement techniques, and type of instrumentation should be used at both the survey unit and the reference area. The distribution of background measurements in the reference area should be the same as that which would be expected in the survey unit if that survey unit had never been contaminated. It may be necessary to select more than one reference area for a specific site, if the site includes so much physical, chemical, radiological, or biological variability that it cannot be represented by a single reference background area.

2.2.8 Radionuclide-Specific Measurements

As indicated in Section 2.2.5 and 2.2.6, if radionuclide-specific survey methods are used, and if the radionuclide of interest does not appear in background, reference area measurements are not needed, and one-sample statistical tests are used. If other survey methods are used, such as gross activity or exposure rate measurements, then the individual contributions due to background and any residual radioactivity will not be separately identifiable, suitable reference area measurements will be needed, and two-sample statistical tests are used.

Even if the radionuclide of interest does appear in background, the variability of radionuclide specific measurements will generally be smaller than those of gross activity measurements in the same area. Depending on the level of residual activity that it is necessary to detect, many more measurements may be required if gross activity or exposure rate measurements are used than if radionuclide-specific measurements are made. At very low levels, it may be difficult or impossible to distinguish the residual radioactivity contribution unless radionuclide-specific methods are used. However, it may be economical in some circumstances to perform a larger number of simpler, less expensive measurements. One of the primary advantages of the Data Quality Objectives process, is that alternative measurement strategies can be compared at the planning stage. Exploring the statistical design of the final status survey in advance, the most efficient method for the problem can be chosen.

2.3 Statistical Concepts

This section introduces some of the statistical concepts and terminology used in hypothesis testing. A use of statistical hypothesis testing that is familiar in the radiation protection measurements field is the calculation of lower limits of detection. The methodology of this report can be viewed as an application of these same concepts to a survey unit rather than to a laboratory measurement. This analogy is pursued further in Section 2.6.

2.3.1 Null and Alternative Hypotheses

The decisions necessary to determine compliance with the criteria for license termination are formulated into precise statistical statements called hypotheses. The truth of these hypotheses can be tested with the survey unit data. The state that is presumed to exist in reality is expressed as the null hypothesis (denoted by $\mathbf{H_0}$). For a given null hypothesis, there is a specified alternative hypothesis (denoted as $\mathbf{H_a}$), which is an expression of what is believed to be the state of reality if the null hypothesis is not true.

For the purposes of this report, the important decision is whether or not a site meets the applicable license termination and release criteria. This decision will be supported by the individual decisions on whether each survey unit meets the applicable release criteria. In this report, two different scenarios, designated Scenario A and Scenario B, are considered.

In Scenario A, the null hypothesis is:
H_0: The survey unit does not meet the release criterion
versus the alternative
H_a: The survey unit meets the release criterion.

In Scenario B, the null hypothesis is:
H_0: The survey unit meets the release criterion.
versus the alternative
H_a: The survey unit does not meet the release criterion.

As indicated in Section 2.2.1, the release criterion is specified in terms of a dose, which is converted via pathway modeling to a residual radioactivity concentration limit, the $DCGL_W$. If the concentration of residual radioactivity that is distinguishable from background in the survey unit exceeds the $DCGL_W$, the survey unit does not meet the release criterion.

When choosing the scenario to use, it is important to note that the null hypothesis cannot be proved, i.e. accepted as true. The null hypothesis is either rejected or not rejected. The data are either consistent with the null hypothesis, or they are not. It is stated this way because there are two circumstances leading to the decision not to reject the null hypothesis:
(1) the null hypothesis is true.
(2) the null hypothesis is false, but the data did not provide enough evidence to show it.

The burden of proof is on the alternative. Thus, in Scenario A, the survey unit will not be released until proven clean. In Scenario B, the survey unit will be released unless it is shown to be contaminated above background. Rejecting the null hypothesis has different implications for survey unit release in the two scenarios. For this reason, a survey unit will be said to *pass* the final status survey if it is concluded that it may be released. Otherwise it will be said to *fail*. In Scenario A, the emphasis is on the dose limit. In Scenario B, the emphasis is on indistinguishability from background. In Scenario A, the survey unit is assumed to fail unless the data show it may be released. In Scenario B, the survey unit is assumed to pass unless the data show that further remediation is necessary.

In Scenario A, the measured average concentration in the survey unit must be significantly less than the $DCGL_W$ in order to pass. In Scenario B, the measured average concentration in the survey unit must be significantly greater than background in order to fail. In Scenario A, increasing the number of measurements in a survey unit increases the probability that an adequately remediated survey unit will pass. In Scenario B, increasing the number of measurements in a survey unit increases the probability that an inadequately remediated survey unit will fail.

Which scenario should be used? Because of insufficient evidence, the null hypothesis may not be rejected even when it is false. Thus, the null hypothesis should be the one that is the easiest to live with even if it is false. The alternative should be the hypothesis that carries the severest consequences if it falsely chosen. To make the proper choice of scenario, the possible types of decision errors and the probability of making them should be examined. This is the subject of the next section.

In most cases, when the $DCGL_W$ is fairly large compared to the measurement variability, Scenario A should be chosen. This is because even contamination below the $DCGL_W$ should be measurable. Requiring additional remediation when it is not strictly necessary may still have some benefit in the form of reduced radiation exposure. Releasing a survey unit that really should

be remediated further is a less tolerable mistake. It is anticipated that Scenario A will be simpler to implement for most licensees.

When the $DCGL_W$ is small compared to measurement and/or background variability, Scenario B should be chosen. This is because contamination below the $DCGL_W$ will be difficult to measure. Requiring additional remediation when it is not necessary, may essentially require remediation of background. This is an impossible task. Releasing a survey unit that has residual radioactivity within the range of background variations is a less severe consequence in this case. It is fairly straightforward to specify what is meant for a survey unit to meet the release criterion, but a survey unit may be distinguishable from background either because it is uniformly contaminated or because it contains spotty areas of residual radioactivity. For this reason, the data analysis for Scenario B involves two statistical tests performed in tandem.

In this report, the two scenarios are developed in parallel. Within the limits imposed by the magnitude of the data variability relative to the $DCGL_W$, essentially the same information about the survey unit should be obtained, and the same conclusion regarding compliance should be reached using either scenario. The difference is in the emphasis.

2.3.2 Decision Errors

Errors can be made when making site remediation decisions. The use of statistical methods allows for controlling the probability of making decision errors. When designing a statistical test, acceptable error rates for incorrectly determining that a site meets or does not meet the applicable decommissioning criteria must be specified. In determining these error rates, consideration should be given to the number of sample data points that are necessary to achieve them. Lower error rates require more measurements, but result in statistical tests of greater power and higher levels of confidence in the decisions. In setting error rates, it is important to balance the *consequences* of making a decision error against the *cost* of achieving greater certainty.

There are two types of decision errors that can be made when performing the statistical tests described in this report. The first type of decision error, called a Type I error, occurs when the null hypothesis is rejected when it is actually true. A Type I error is sometimes called a "false positive." The probability of a Type I error is usually denoted by α. The Type I error rate is often referred to as the significance level or size of the test.

The second type of decision error, called a Type II error, occurs when the null hypothesis is not rejected when it is actually false. A Type II error is sometimes called a "false negative." The probability of a Type II error is usually denoted by β. The *power* of a statistical test is defined as the probability of rejecting the null hypotheses when it is false. It is numerically equal to $1-\beta$, where β is the Type II error rate.

The setting of acceptable error rates is a crucial step in the planning process. Specific considerations for establishing these error rates are discussed in Chapter 3. Table 2.2 summarizes the types of decision errors that can be made for the specific hypotheses of Scenario A and Scenario B.

Table 2.2 Summary of Types of Decision Errors

Scenario A	**True Condition of Survey Unit**	
Decision Based on Survey	**Does Not Meet Release Criterion**	**Meets Release Criterion**
Does Not Meet Release Criterion	*Survey unit fails* Correct Decision (Probability = 1-α)	*Survey unit fails* Type II Error (Probability = β)
Meets Release Criterion	*Survey unit passes* Type I Error (Probability = α)	*Survey unit passes* Correct Decision (Power = 1-β)

Scenario B	**True Condition of Survey Unit**	
Decision Based on Survey	**Meets Release Criterion**	**Does Not Meet Release Criterion**
Meets Release Criterion	*Survey unit passes* Correct Decision (Probability = 1-α)	*Survey unit passes* Type II Error (Probability = β)
Does Not Meet Release Criterion	*Survey unit fails* Type I Error (Probability = α)	*Survey unit fails* Correct Decision (Power = 1-β)

2.4 Hypothesis Testing Example

The following example illustrates the use of the concepts discussed above as currently used in the determination of detection limits for radioactivity measurements. The analogy is most direct for Scenario B.[(1)] The calculation of detection limits, which is generally familiar to radiation protection professionals, also involves hypothesis testing (HPSR/ EPA 520/1-80-012, 1980; NUREG/CR-4007, 1984; Currie, 1968). In this situation, there is a measurement error, often taken to be the Poisson counting error, σ, equal to the square root of the number of counts. There is a background counting rate, and any additional radioactivity in a sample must be distinguishable above that. Generally it is assumed that the number of counts is sufficiently large so that a normal approximation to the Poisson distribution of counts is appropriate.

[(1)] For Scenario A, the analogy would have to be restructured for the problem of deciding whether a given sample, assumed to contain added radioactivity exceeding L_D, actually contained a smaller amount. In essence, the null and alternative hypotheses would be reversed.

2.4.1 Detection Limits

For the calculation of detection limits, the hypotheses are:

Null Hypothesis**:**
H_0: The sample contains no radioactivity above background.
and
Alternative Hypothesis:
H_a: The sample contains added radioactivity.

The count obtained from the sample measurement is the test statistic, and it has a different probability distribution under the null and alternative hypothesis (see Figure 2.1). If a sample that contains no radioactivity above background is declared to contain radioactivity above background, a Type I error is made. Conversely, if a sample that contains radioactivity above background is declared to contain no radioactivity above background, a Type II error is made.

The Type I error rate, α, depends on the variability of background, i.e., it is controlled by requiring that the net counts exceed a certain multiple of the measurement standard deviation. Under the null hypothesis, namely when there is no radioactivity above background, the net counts have mean $B - B = 0$.

The standard deviation of the net count is

$$\sigma_{B-B} = \sqrt{B + B} = \sqrt{\sigma^2 + \sigma^2} = \sqrt{2}\,\sigma \qquad (2\text{-}1)$$

where B is the background count, and $\sigma = \sqrt{B}$ is its standard deviation, since for a Poisson distribution the standard deviation is the square root of the mean. Unless the mean number of counts is very low, a normal distribution with the same mean and standard deviation can be used to approximate the Poisson distribution of the background counts. This determines the critical level, L_C. If a net count above the critical detection level is obtained, the null hypothesis is rejected. That is, the decision is made that the sample being measured contains radioactivity above background.

$$L_C = z_{1-\alpha}\,\sigma_{B-B} = z_{1-\alpha}\,\sqrt{2}\,\sigma \qquad (2\text{-}2)$$

$z_{1-\alpha}$ is the $1-\alpha$ percentile of a standard normal distribution, e.g. if $\alpha = 0.05$, then $1 - \alpha = 0.95$ and $z_{1-\alpha} = 1.645$. Note that the distribution of background counts (lefthand curve in Figure 2.1) is used for this calculation.

The Type II error rate, β, depends on the variability of the added radioactivity and is controlled by requiring that the net counts exceed a certain number of standard deviations above the critical level.

$$L_D = L_C + z_{1-\beta}\,\sigma_{L_D} = z_{1-\alpha}\,\sqrt{2}\,\sigma + z_{1-\beta}\,\sigma_{L_D} = z_{1-\alpha}\,\sqrt{2}\,\sigma + z_{1-\beta}\,\sqrt{L_D + 2\sigma^2} \qquad (2\text{-}3)$$

since

$$\sigma_{L_D} = \sqrt{(L_D + B) + B} = \sqrt{L_D + 2\sigma^2}$$

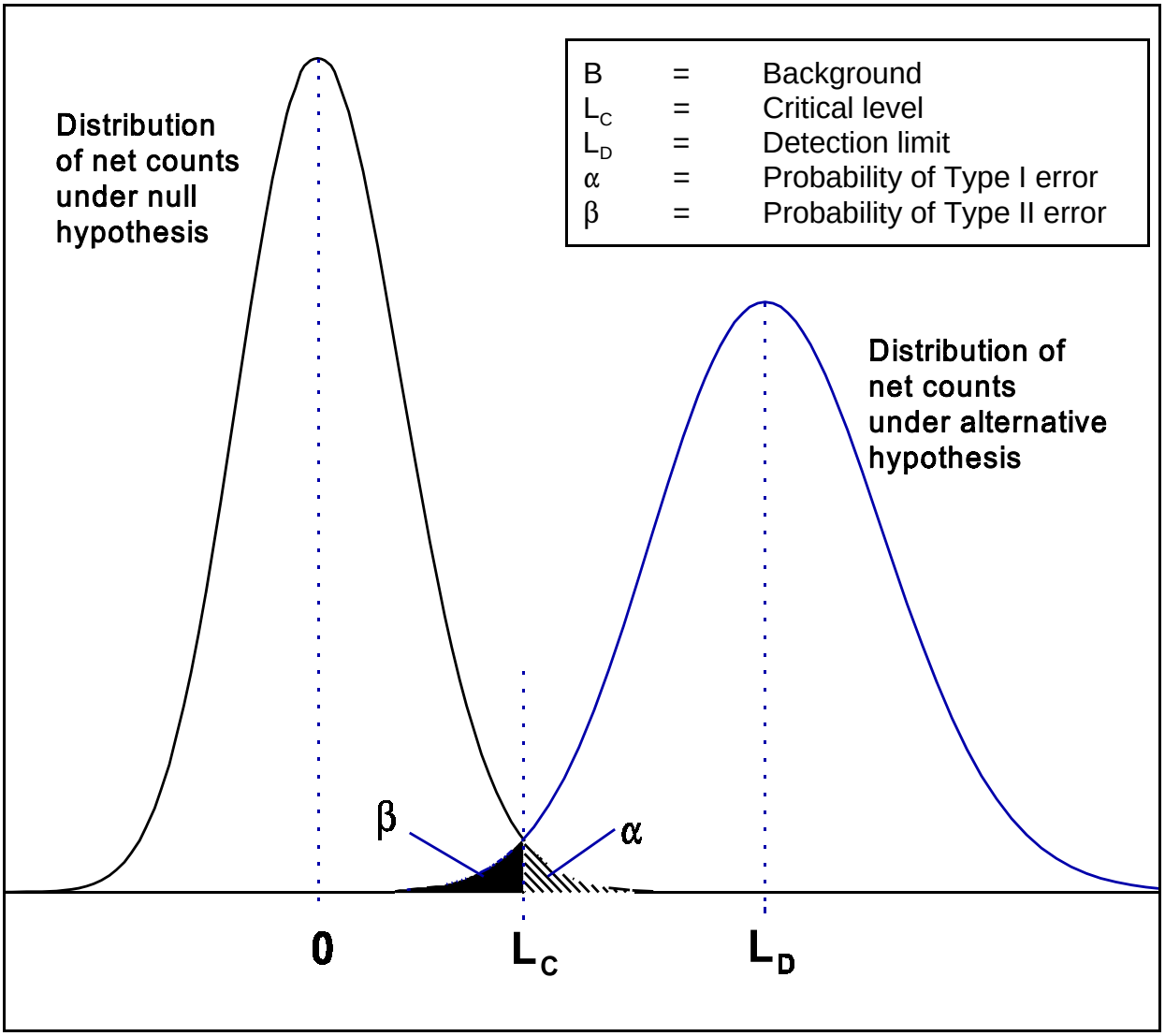

Figure 2.1 Type I and Type II Errors in the Determination of a Detection Limit

The distribution of counts under the alternative hypothesis (right hand curve in Figure 2.1) is used to derive Equation 2-3. If the probability of a Type II error is set the same as the probability of a Type I error, then $z_{1-\alpha} = z_{1-\beta} = k$. Solving Equation 2-3 for L_D, the count detection limit is found to be

$$L_D = k^2 + 2k\sqrt{2}\,\sigma = k^2 + 2\,L_C \qquad (2\text{-}4)$$

The power, $1-\beta$, is the probability that the measurement will indicate the presence of additional radioactivity in the sample, when the sample actually contains additional activity in the amount necessary to produce an average of L_D counts above background during the measurement.

2.4.2 Final Status Surveys

The statistical procedures described in this report for final status surveys have many similarities to the detection limit calculation. Corresponding to Figure 2.1, the relationship between the hypothesis, decision error rates and measurement distributions in Scenario A and Scenario B are shown in Figures 2.2 and 2.3, respectively.

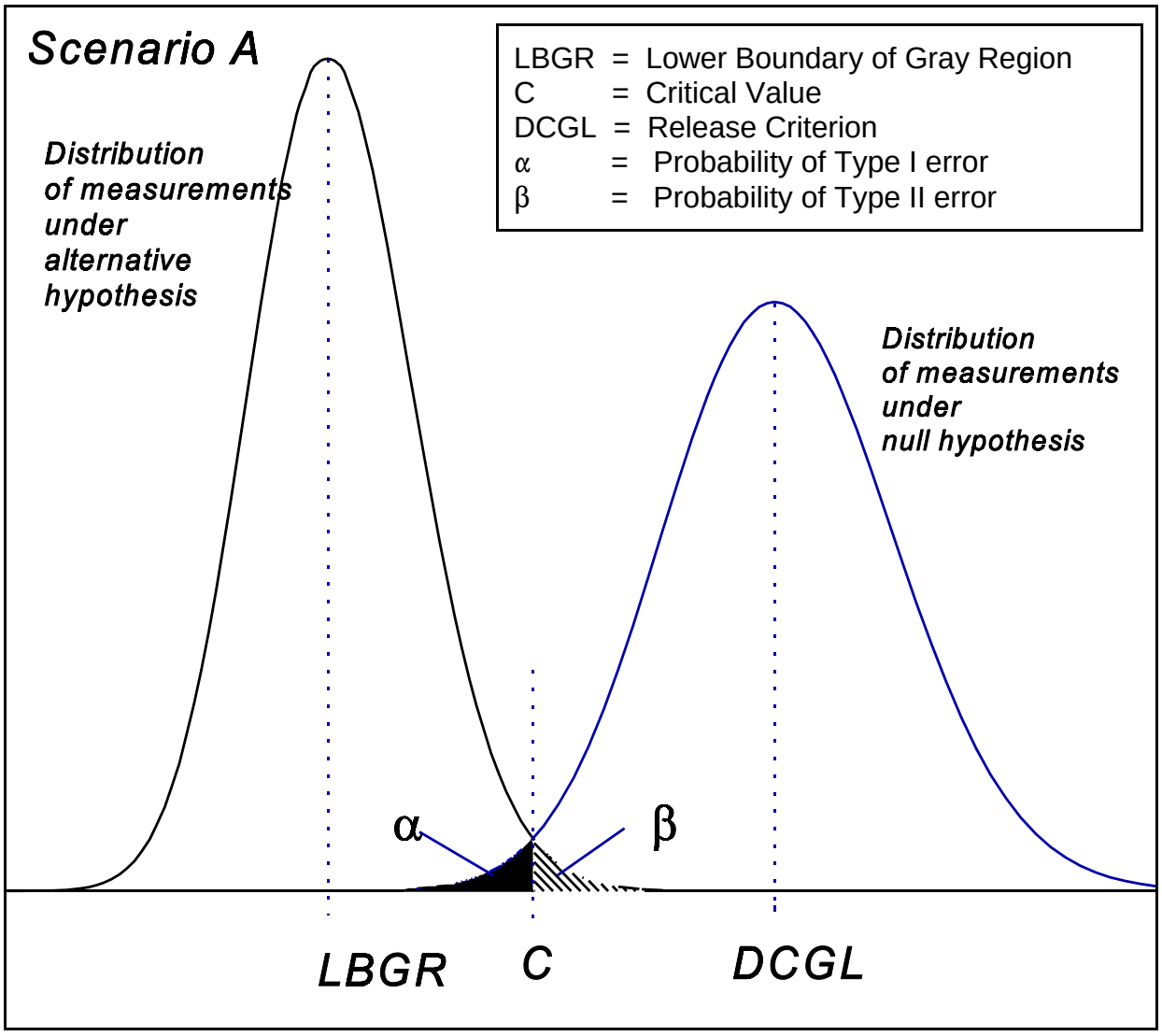

Figure 2.2 Type I and Type II Errors for Scenario A

Some other points of similarity are:

(1) The null hypothesis is:

H_0: The sample contains no radioactivity above background.

becomes either

H_0: The survey unit does not meet the release criterion (Scenario A).

or

H_0: The survey unit meets the release criterion (Scenario B).

(2) The alternative hypothesis is:

H_a: The sample contains added radioactivity above the detection limit.

becomes either

H_a: The survey unit meets the release criterion (Scenario A).

or

H_a: The survey unit does not meet the release criterion (Scenario B).

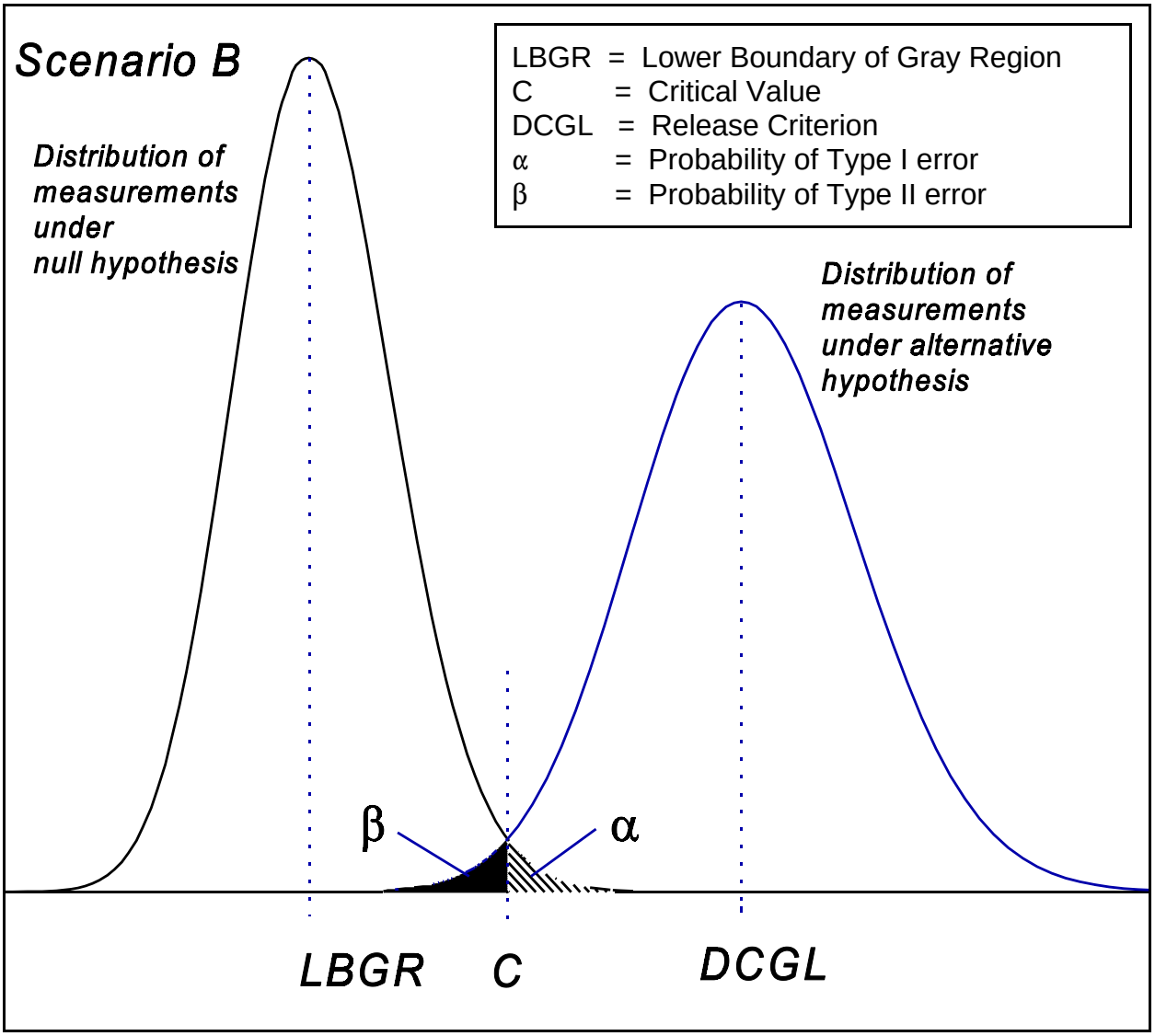

Figure 2.3 Type I and Type II Errors for Scenario B

(3) The Type I error rate is computed using the distribution of counts assuming the null hypothesis is true. Similarly, the Type I error rates for the tests described in this report will be calculated using the distribution of the measurements under the null hypothesis.

(4) The Type II error rate is computed using the distribution of counts assuming the alternative hypothesis is true. Similarly, the Type II error rates for the tests described in this report will

be calculated using the distribution of the measurements under the alternative hypothesis. This also gives the power of the tests.

(5) The variability of the count obtained from the sample, σ, plays a crucial role in determining the value of the detection limit. Similarly, the variability of the radioactivity measurements in the reference areas and survey units plays a crucial role in how well the tests described in this report will perform.

(6) Corresponding to the detection limit , L_D , is a critical level of counts, L_C . Any sample producing more than the critical level of counts is assumed to contain additional radioactivity. Thus, the decision whether or not to reject the null hypothesis is based on comparing the counts actually obtained from the sample to the critical detection level. Similarly, the decision whether or not to reject the null hypothesis for a survey unit is based on the critical level of a test statistic which is computed from the measurement data. Note that while L_C and L_D are expressed in counts, there is a corresponding concentration level in the sample being measured that will, on average, give rise to that number of counts.

(7) The critical level of counts, $L_{C,}$ is calculated so that the decision to reject the null hypothesis is made with probability α when the true concentration in the sample being measured is zero. The critical value of the final status survey test statistic is calculated so that the decision to reject the null hypothesis in Scenario B is made with probability α when the true concentration is equal to a certain value called the LBGR (Lower Boundary of the Gray Region). The LBGR is a concentration value between zero and the $DCGL_W$ at which probability of the survey unit incorrectly failing the final status survey is specified. The LBGR is discussed further in Section 3.7.

(8) The detection limit can usually be made lower by counting for a longer time, thereby reducing the relative measurement error, at additional cost. Similarly, the ability of the tests described in this report to distinguish smaller amounts of residual radioactivity from background more accurately can be improved by taking a greater number of samples, at additional cost.

(9) Usually, a detection limit is calculated given the Type I and Type II error rates and the background variability. However, if a certain detection limit is pre-specified instead, the procedure given above shows how to relate it to the Type I and Type II error rates, and the measurement variability. Similarly, the procedures of this report will show the interrelationship of the decommissioning criteria (dose above background), the Type I and Type II error rates, and the measurement variability.

2.4.3 The Effect of Measurement Variability on the Decisionmaking Process

Figure 2.4 further illustrates the affect of the measurement standard deviation on the decision process. Shown are three hypothetical measurement distributions, with true mean concentration equal to zero, one, and three times the measurement standard deviation, σ. Assume for simplicity there is no background to subtract. Then the critical level, $L_C = z_{1-\alpha}\sigma = 1.645\sigma$ when $\alpha = 0.05$. Thus, there is a 5% chance of a positive result when the true concentration is actually zero. If the true concentration is 3σ, the probability of a positive result is very high since most of the distribution lies above L_C (91% using the normal distribution table with $z = 3 - 1.645 = 1.355$).

However, if the true concentration is 1σ, then there is less than a 50% chance of a positive result (26% using the normal distribution table with $z = 1 - 1.645 = -0.645$). If a true mean concentration of $C = 1\sigma$ must be measured, then the uncertainty must be reduced by taking more measurements. If nine measurements are averaged, then the standard deviation of the mean, σ*, falls by a factor of three (one over the square root of the number of measurements). In the "new standard deviation units" $C = 1\sigma = 3\sigma^*$. Thus, a difference of 1σ can be distinguished with nine measurements as easily as a difference of 3σ can be distinguished with one measurement.

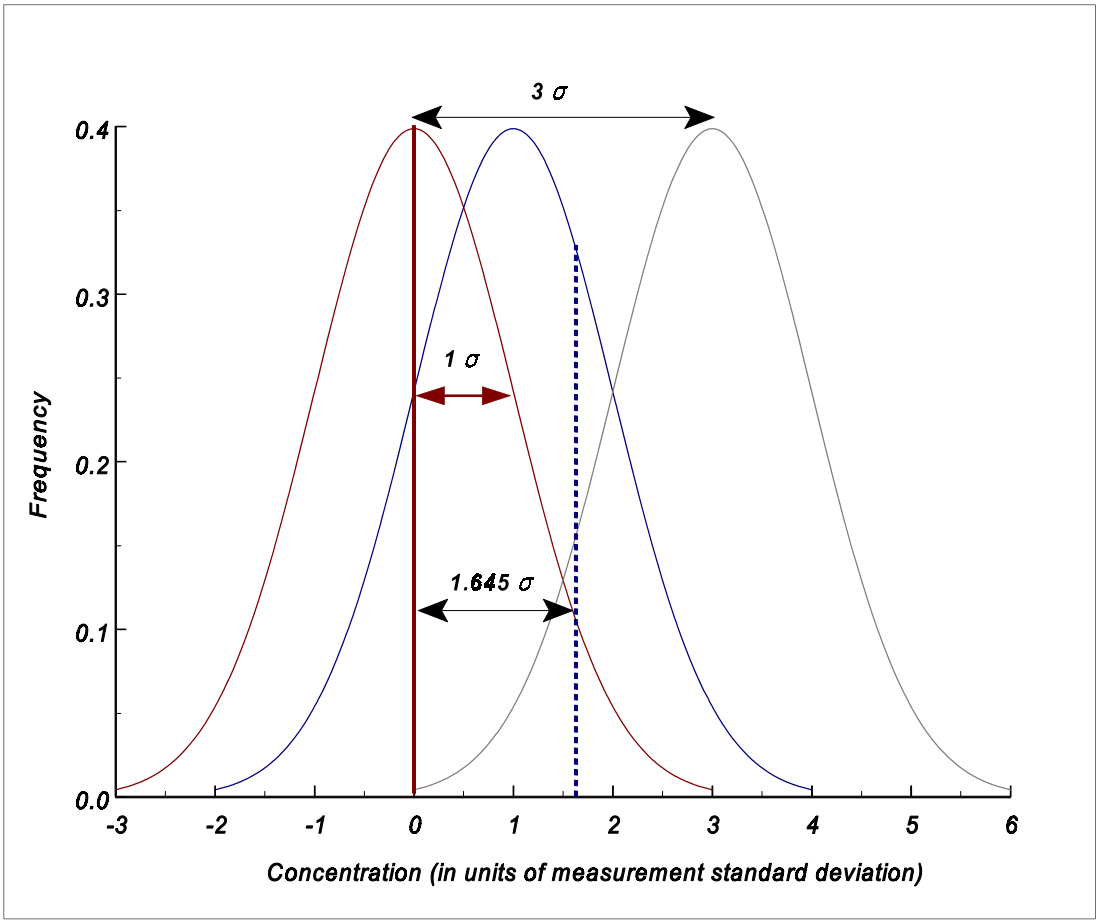

Figure 2.4 Differences in Concentration Compared to Measurement Variability

2.5 Statistical Tests

There are two important uses of the statistical tests described in this report. The first is in the analysis of the final status survey data to demonstrate compliance with the release criterion. However, the second, and perhaps more important use, is in the design of the final status survey. In some cases it may be clear from the data, without any formal analysis, whether or not a survey unit meets the decommissioning criteria. Provided that an adequate number of measurements are made (either *in situ* or from samples), Table 2.3 can be used to determine whether or not a formal statistical test is necessary.

It is usually not obvious, *a priori*, what number of samples is necessary in order to show whether or not a survey unit meets the decommissioning criteria with acceptable rates of error. The DQO process described in Chapter 3 provides a general method for designing surveys so that accurate remediation decisions can be made cost effectively. Simple estimates of the number of samples required for the statistical tests may be made using the mathematical relationships between the error rates, residual radioactivity levels, and measurement variability.

Table 2.3 Summary of Statistical Tests

Radionuclide not in background *and* radionuclide-specific measurements made	
Survey Result	**Conclusion**
All measurements below $DCGL_W$	Survey unit meets release criterion
Average above $DCGL_W$	Survey unit does not meet release criterion
Otherwise (some measurements above $DCGL_W$, but average below $DCGL_W$)	Conduct Sign test and elevated measurement comparison

Radionuclide in background *or* non-radionuclide-specific measurements made	
Survey Result	**Conclusion**
Difference between maximum survey unit measurement and minimum reference area measurements is below $DCGL_W$	Survey unit meets release criterion
Difference of survey unit average and reference area average is above $DCGL_W$	Survey unit does not meet release criterion
Otherwise (Maximum difference above $DCGL_W$, but average difference below $DCGL_W$)	Conduct WRS test and elevated measurement comparison

2.5.1 Nonparametric Statistical Tests

Many statistical tests can be used for determining whether or not a survey unit meets the release criteria. Any one test may perform better or worse than others, depending on the hypotheses to be tested, i.e., the decision that is to be made and the alternative, and how well the assumptions of the test fit the situation.

The basic distinction between parametric and nonparametric statistical techniques is that parametric techniques use specific assumptions about the probability distributions of the measurement data. The most commonly made assumption is that the data fit a normal distribution. Such is the case when the Student's t statistic is used. Additional data and statistical tests would generally be necessary in order to show that the assumption of normality is justified (EPA QA/G-9), 1995.

Nonparametric techniques (sometimes referred to as distribution-free statistical methods) can be used without assuming a particular underlying distribution. Thus, nonparametric techniques are often more appropriate in situations when the probability distribution of the data is either unknown or is some continuous distribution other than the normal distribution. That a statistical approach is nonparametric or distribution free does not imply that it is free of any and all assumptions about the data distribution. Most nonparametric procedures require that measured values be independent and identically distributed. Some nonparametric procedures assume that the underlying probability distribution is symmetric. However, these requirements are usually less restrictive than the assumption that the data follow a particular symmetric distribution, such as the normal distribution.

Parametric methods rely on the assumptions about the data distribution to infer how large the difference between two measurements is expected to be. These methods are better only if the assumptions are true. Many nonparametric techniques are based on ranking the measurement data. The data are ordered from smallest to largest, and assigned numbers (ranks) 1, 2, 3,... accordingly. The analysis is then performed on the *ranks* rather than on the original measurement values. The advantage of this approach is that the probability that one measurement is larger (i.e. ranked higher) than another can be computed exactly by combinatorial (enumeration and counting) methods without reference to a specific probability distribution.

If the underlying distribution is known, a parametric test can make use of that additional information. If the underlying distribution is different than that assumed, however, the results can be unpredictable. The nonparametric methods described in this report have been found to perform nearly as well as the corresponding parametric tests, even when the conditions necessary for applying the parametric tests are fulfilled. There is often relatively little to be gained in efficiency from using a specific parametric procedure, but potentially much to be lost. Thus, it may be considered prudent to use nonparametric methods in all cases.

For survey measurements at or near background, there may be some measurement data which are at or below instrumental detection limits. These data are sometimes reported as “less than” or “non-detects”. Such data are not easily treated using parametric methods. It is recommended that the actual numerical results of measurements always be reported, even if these are negative or below calculated detection limits. However, if it is necessary to analyze data which include non-numerical results, nonparametric procedures based on ranks can still be used in many cases. This is an additional advantage to the use of these methods.

2.5.2 Wilcoxon Rank Sum and Sign Tests

The Wilcoxon Rank Sum (WRS) test and Sign test are used in this report to detect a uniform shift in the mean of a distribution of measurements. Without assuming symmetry in the measurement distribution, these tests are technically for the median. However, computer simulations have shown that these tests generally produce the correct decisions more often when the assumption of symmetry is violated than the commonly used Student’s t-test, which assumes normality in addition to symmetry. Nevertheless, extremes of asymmetry are guarded against by conducting the elevated measurement comparison in addition to the WRS and Sign tests. This issue is discussed further in the next section.

The WRS test is a two-sample test that compares the distribution of a set of measurements in a survey unit to that of a set of measurements in a reference area. The Sign test is a one-sample test that compares the distribution of a set of measurements in a survey unit to a fixed value, namely the derived concentration limit for a specific radionuclide.

The WRS test, also known as the Mann-Whitney test, is performed for Scenario A by first adding the value of the DCGL to each of the reference area measurements, and then listing the *combined* set of survey unit and adjusted reference area measurements in increasing numerical order from smallest to largest. The next step is to replace the measurement values by their ranks, i.e., their position number in the ordered list. Thus, the ranks are simply integer values from 1 through N, where N is the total number of combined measurements. The rank 1 is assigned to the smallest value, 2 to the second smallest observation, etc. Then, the sum of the ranks of the survey unit and the sum of the ranks of the adjusted reference area measurements is computed. Because the sum of the combined ranks is a fixed constant equal to $N(N+1)/2$, the sum of the adjusted reference area measurement ranks is equal to $N(N+1)/2$ minus the sum of the ranks of the survey unit measurements.

If the level of residual radioactivity in the survey unit is exactly at the DCGL, then any given rank is equally likely to belong to either an adjusted reference area measurement or a survey unit measurement. If the survey unit has residual radioactivity less than the DCGL, the survey unit site ranks will tend to be smaller than the adjusted reference area ranks. The larger the average of the ranks of the adjusted reference area measurements relative to the ranks of the survey unit measurements, the smaller the probability that it is by chance, and the greater the evidence that the residual radioactivity in the survey unit is actually below the DCGL. If the sum of the ranks of the adjusted reference area measurements exceeds a calculated critical value, the decision is made to reject the null hypothesis and to conclude that the survey unit actually meets the release criterion. In some cases, the result will be obvious without any computations. If, for example, all of the survey unit measurements are less than the smallest of the adjusted reference area measurements, then the sum of the ranks of the adjusted reference area measurements is at its maximum possible value, and the null hypothesis will always be rejected.

For Scenario B, the WRS is performed by first subtracting the value of the LBGR from each of the survey unit measurements, and then listing the combined set of adjusted survey unit and reference area measurements in increasing numerical order from smallest to largest and finding their ranks. As above, the sum of the ranks of the survey unit measurements and the sum of the ranks of the unadjusted reference area measurements is computed. Again, the sum of the reference area measurement ranks plus the sum of the ranks of the adjusted survey unit measurements is equal to $N(N+1)/2$. If the level of residual radioactivity in the survey unit is exactly at the LBGR, then any given rank is equally likely to belong to either a reference area measurement or a survey unit measurement. Thus, there is no reason to believe that the average of the survey unit ranks will differ greatly from the average of the reference area ranks. With residual radioactivity at the LBGR, the probability that the average of the survey unit ranks will be larger than the average of the reference area ranks is 50% by random chance. However, the larger the average of the survey unit ranks, the smaller the probability that it is by chance, and the greater the evidence that the survey unit is contaminated. If the sum of the survey unit ranks exceeds a calculated critical value, one can decide that the evidence shows that the residual radioactivity in the survey unit exceeds the LBGR.

The one-sample Sign test is performed for Scenario A by first subtracting each survey unit measurement from the derived concentration limit. Then the *number* of *positive* differences is counted. Large numbers of positive differences are evidence that the survey unit measurements do not exceed the derived concentration guideline.

The Sign test is performed for Scenario B by first subtracting the LBGR from each survey unit measurement. Then the *number* of *positive* differences is counted. Large numbers of positive differences are evidence that the survey unit measurements exceed the LBGR.

The Sign test uses no assumptions about the shape of the data distribution. An alternative test, the Wilcoxon Signed Rank (WSR) test, assumes that the measurement distribution is symmetric. When this assumption is valid, the WSR test generally has greater power than the Sign test. The WSR and other alternative statistical tests are discussed in Section 14.1.

2.5.3 Mean and Median

The WRS and Sign tests are designed to determine whether or not a degree of residual radioactivity remains uniformly throughout the survey unit. Since these methods are based on ranks, the results are generally expressed in terms of the median. When the underlying measurement distribution is symmetric, the mean is equal to the median. The assumption of symmetry is less restrictive than that of normality, since the normal distribution is itself symmetric. If, however, the measurement distribution is skewed to the right, the average will generally be greater than the median. In severe cases, it may happen that the average exceeds the $DCGL_W$ while the median does not. This is why the average is used to screen the data set before any statistical test is performed (see Table 2.3).

Figure 2.5 illustrates the potential differences between the median and the mean. The normal and lognormal distributions are commonly used examples of symmetric and skewed distributions, respectively. In this figure, the probability densities all have arithmetic mean equal to one. The coefficient of variation (arithmetic standard deviation divided by the mean) varies from 0.1 to 1.0. For values of the coefficient of variation larger than about 0.3, the lognormal begins to diverge significantly from the normal. When the coefficient of variation is 1.0, the difference between the median and the mean is large.

When the underlying data distribution is highly skewed, it is often because there are a few high measurements. Since the elevated measurement comparison is used to detect such measurements, the difference between using the median and the mean as a measure for the degree to which uniform residual radioactivity remains in a survey unit tends to diminish in importance. This is especially true in Scenario A, where the null hypothesis is that the survey unit does not meet the release criterion.

In Scenario B, the null hypothesis is that the survey unit meets the release criterion. If the measurement distribution is highly skewed, there may a substantial portion of the survey unit with residual radioactivity higher than the $DCGL_W$, but perhaps not in excess of the $DCGL_{EMC}$. In such cases, the median may be below the LBGR, while the mean is above it. The *Quantile* test, discussed in Section 2.5.5, can be used to detect when remediation activities have failed in only a few areas within a survey unit. Conducting the Quantile test in tandem with the WRS has been

found to improve the accuracy of the tests (EPA 230-R-94-004, 1994).

In some cases, data from an asymmetric distribution can be transformed so that the transformed data have a more symmetric distribution. The analysis is then performed using the transformed data. A common example is that the logarithms of lognormally distributed data have a normal distribution. However, such transformations introduce additional complications. The mean of the transformed data is not generally equal to the transform of the mean of the original data. For example, the mean of the logs transforms back to the geometric mean, which is the median of a lognormal data set. The computations necessary for testing the average of lognormal data can be complex (see Chapter 14). The behavior of this lognormal test when the assumption of lognormality is violated is not known.

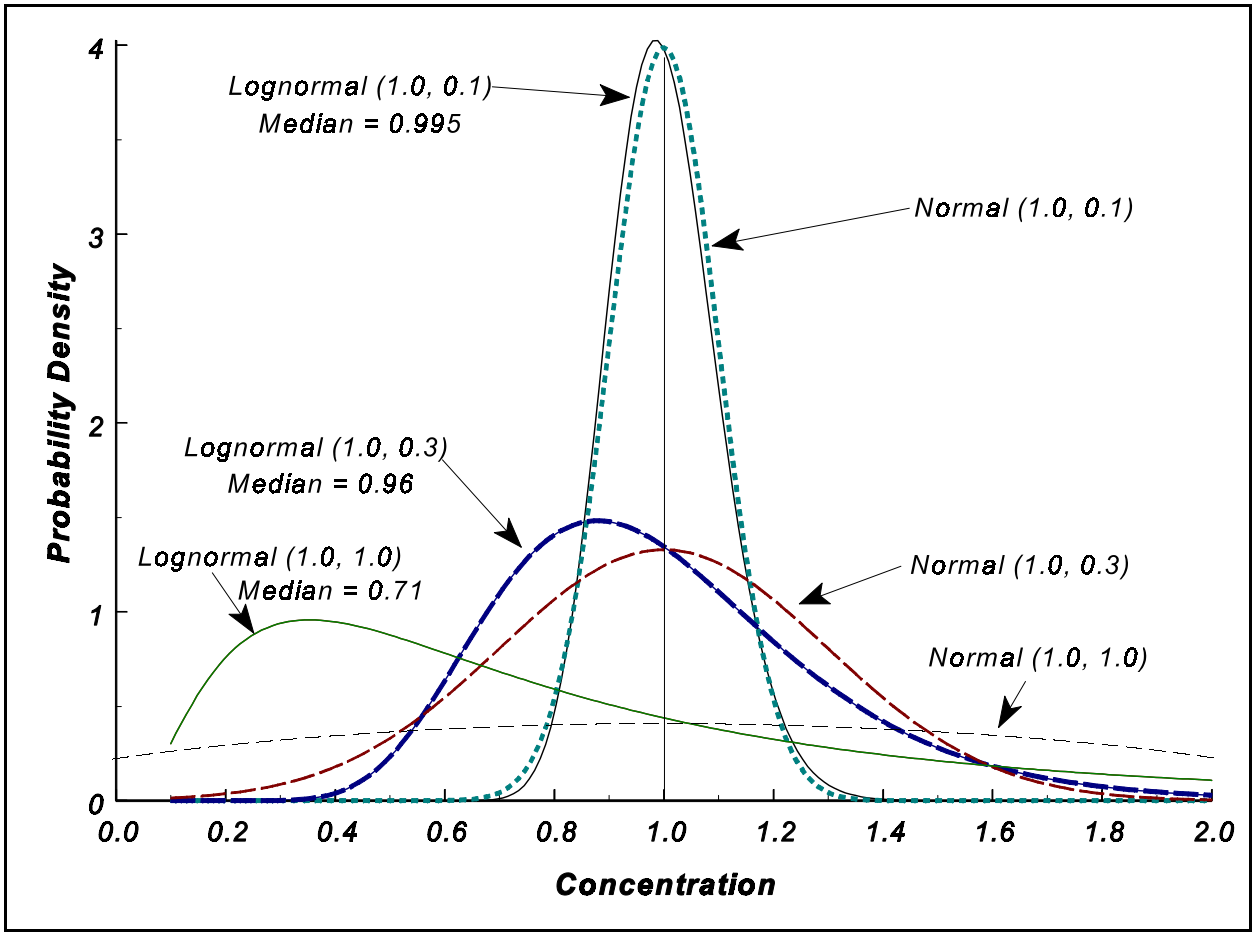

Figure 2.5 Comparison of Normal and Lognormal Distributions
(Arithmetic mean and arithmetic standard deviation shown in parentheses)

The EPA (EPA QA/G-4, 1994) has compared the use of the mean and the median as the parameter of interest whose true value the decision maker would like know and that the data will estimate. Some of the positive and negative attributes of each are listed below:

MEAN

Positive Attributes

- Useful when action level is based on long-term, average health effects
- Useful when the population is uniform with relatively small spread.
- Generally requires fewer samples than other parameters.

Negative Attributes

- Not a very representative measure of central tendency for highly skewed populations.
- Not useful when the population contains a large proportion of values that are less than measurement detection limits.

MEDIAN

Positive Attributes

- Useful when action level is based on long-term, average health effects
- More representative measure of central tendency than the mean for skewed populations.
- Useful when the population contains a large number of values that are less than measurement detection limits.
- Relies on few statistical assumptions.

Negative Attributes

- Will not protect against the effect of extreme values.
- Not a very representative measure of central tendency for highly skewed populations.

2.5.4 Estimating the Amount of Residual Radioactivity

The result of the statistical test is the decision whether or not to reject the null hypothesis. Following the statistical hypothesis tests, it may also be necessary to estimate the amount of residual radioactivity in the survey unit so that dose calculations can be made. It is usually best to use the mean (average) residual radioactivity for this purpose (EPA PB92-963373, 1992). If the data distribution is symmetric, the mean is equal to the median. If, however, the data distribution is skewed, the mean may be greater than the median.

2.5.5 Quantile Test

In contrast to locations with concentrations above the $DCGL_{EMC}$, more moderate departures from uniformity in residual radioactivity concentrations may also exist within a survey unit. One portion of a survey unit may have virtually no residual radioactivity, while another portion does contain some residual radioactivity. There may be several portions of one type or another in a survey unit, resulting in a patchy contamination pattern. The Quantile test is designed to detect this type of residual radioactivity. The Quantile test is only needed in Scenario B.

Like the WRS test, the Quantile test (EPA 230-R-94-004, 1994; Johnson et al., 1987) is a two-sample test. It is also performed by first subtracting the value of the LBGR from each of the survey unit measurements, and then listing the combined set of adjusted survey unit and reference area measurements from smallest to largest. However, only the largest measurements in the list are examined. The number of measurements that will be considered in the Quantile test is denoted by *r*. A count is made of the number of measurements among the largest *r* measurements that are from the survey unit. This number is denoted by *k*. If there is no contamination, measurements from the reference area and from the survey unit would be

expected to appear among the *r* largest measurements roughly in proportion to the number of measurements made in each of them. If patchy residual contamination exists, then the *r* largest measurements of the combined data sets (reference area and survey unit) are more likely to come from the survey unit. When there are *m* background measurements and *n* survey site measurements, then *k* should be about *r* times *n/(m+n)*. If the number of measurements from the survey unit among the largest *r* is too much larger than this, then there is evidence that the survey unit contains patchy areas of residual radioactivity in excess of the LBGR.

While it is possible to perform a one-sample version of the Quantile test, it will seldom be necessary in practice. With no interfering background, patchy areas of contamination in excess of the LBGR will be apparent in simple posting plots and histograms (see Chapter 4) of the survey unit data.

2.5.6 Elevated Measurements Comparison

An *elevated measurement comparison* is performed by comparing each measurement from the survey unit to the $DCGL_{EMC}$. If the survey unit is being compared to a reference area, the net survey unit measurement is first obtained by subtracting the mean of the reference area measurements. A net survey unit measurement that equals or exceeds the $DCGL_{EMC}$ is an indication that a survey unit may contain residual radioactivity in excess of the release criterion.

This type of measurement comparison is sometimes called a "hot spot test." The latter term may be misleading because it is not a formal statistical test, but a simple comparison of measured values against a limit. In addition, there is no commonly accepted definition of what constitutes a *hot spot* in either area or magnitude of residual radioactivity. Yet, this term may imply some degree of radiological hazard. In this report, the term "area of elevated residual radioactivity" is used to describe a limited area of residual activity that may cause the decommissioning dose criteria to be exceeded. It is only these areas that might be considered *hot spots*. For planning purposes, the potential extent of an "area of elevated residual radioactivity" is based on the distance between sampling points in the survey sampling grid.

In addition to direct measurements or samples at discrete locations, parts of each survey unit will also be scanned. For the quantitative measurements obtained at discrete locations, performing the EMC is a straightforward comparison of two numerical values. Some sophisticated scanning instrumentation is also capable providing quantitative results with a quality approaching those from direct measurements or samples. Other scanning measurements, however, may be more qualitative. In that case, *action levels* should be established for the scanning procedure so that areas with concentrations that may exceed the $DCGL_{EMC}$ are marked for a quantitative measurement.

A single unusually large measurement may occur by chance. The elevated measurement comparison flags these measurements for further study. When a measurement is flagged, it should first be determined that it is not due to sampling or analysis error. Such a determination may include re-sampling the area at which the measurement was originally taken.

If an measurement exceeding the is $DCGL_{EMC}$ is confirmed, then the size of the area of elevated residual radioactivity, A', and the average concentration within it, $C_{A'}$, is determined. This will

generally involve taking further measurements in the vicinity of the elevated measurement to adequately delimit its extent. Using the area factor $F_{A'}$ for the area A', $C_{A'}$ should not exceed the product $(F_{A'})(DCGL_W)$ in order for the survey unit to meet the release criterion.

2.5.7 Investigation Levels

In contrast to an elevated measurement, a measurement may be found that exceeds the concentration level expected from the survey unit's classification. *Investigation levels* are established for each class of survey unit to guard against the possible mis-classification of survey units. If a measurement exceeds the investigation level, additional investigation is required to determine if the final status survey for the survey unit was adequate to determine compliance with the release criteria.

For example, in a Class 1 survey unit, measurements above the $DCGL_W$ may not be unusual or unexpected. In Class 2 areas, however, neither measurements above the $DCGL_W$ nor elevated areas are expected. Thus in these areas, any measurement at a discrete location exceeding the $DCGL_W$ should be flagged for further investigation. Unless the scanning sensitivity is such that an action level can be specified for areas with concentrations potentially exceeding the $DCGL_W$, any positive indication of residual radioactivity during the scan could warrant further investigation.

Because there is a low expectation of any residual radioactivity in a Class 3 area, it may be prudent to investigate any measurement exceeding even a fraction of the $DCGL_W$. What level that should be will depend on the site, the radionuclides of concern, and the measurement and scanning methods chosen. This level should be set using the DQO Process during the survey design phase of the Data Life Cycle. In some cases it may be prudent to follow this procedure for Class 2 survey units as well.

Suggested investigation levels that might be appropriate for each class of survey unit and type of measurement are shown in Table 2.4.

Table 2.4 Summary of Investigation Levels

Survey Unit Classification	Flag Direct Measurement or Sample	Scanned Area Marked When Action Levels Indicate:
Class 1	> $DCGL_{EMC}$	>$DCGL_{EMC}$
Class 2	> $DCGL_W$	>$DCGL_W$
Class 3	> fraction of $DCGL_W$	>MDC

In the last two sections, we have considered elevated measurements that require investigation for compliance with the dose criteria, and investigation levels that flag potential survey unit mis-classifications. In addition to these, are the QA/QC procedures that should be in place in any

measurement program. Gross errors of analysis can often be spotted by the use of simple preliminary data analysis techniques, such as posting plots and histograms. This is sometimes called exploratory data analysis. These techniques also form a part of the Data Quality Assessment process (EPA QA/G-9, 1995). Some of these methods are discussed in Chapter 4.